

Pengembangan Fitur Evaluasi dengan Metode Large Language Model = Development of the Evaluation Feature using Large Language Model Method

Fauzan Ramadhanto Andiko Putra, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920575286&lokasi=lokal>

Abstrak

Penelitian ini bertujuan untuk mengembangkan sistem evaluasi yang menggunakan Generative AI berjenis Large Language Model yang dengan output pertanyaan. Sistem yang digunakan menggunakan metode Retrieval-Augmented Generation agar pertanyaan yang dihasilkan memiliki konteks terhadap pertanyaan yang diperintahkan untuk dibuat menggunakan file dokumen yang disediakan oleh pengguna. Selain itu, diterapkan juga prompt engineering untuk mengoptimalkan output dengan memberikan sebuah set instruksi, peran dan contoh. Implementasi fitur evaluasi diterapkan pada TCExam, sebuah platform ujian dan evaluasi open source yang sering digunakan oleh masyarakat. Penelitian ini berhasil dalam mengembangkan sistem evaluasi dengan metode Large Language Model yang terintegrasi dengan platform MOOC sehingga dapat proses penyusunan soal dan pengerjaan soal dapat dilakukan di satu wadah yang sama. Selain itu, evaluasi yang dilakukan pada penelitian ini adalah perbandingan antara dua model yang dapat digunakan pada sistem evaluasi untuk menghasilkan pertanyaan dan evaluasi sistem yang sudah terintegrasi. Dua model yang diuji pada penelitian ini adalah Mistral-7B dan LLaMa-3 8B dengan hasil evaluasi menunjukkan bahwa Mistral-7B menghasilkan output yang lebih baik. Pada evaluasi sistem, penggunaannya cukup mudah dipahami dengan ruang untuk peningkatan yang dapat diterapkan pada penelitian berikutnya.

.....This study aims to develop an evaluation system using Generative AI, specifically Large Language Model with the target output of questions. This system uses Retrieval Augmented Generation to help give context for the questions generated. Furthermore, prompt engineering is implemented to optimize the output by giving a set of instructions, roles, and examples. In implementing the evaluation feature, TCExam is used. It is an open source assessment platform often used by the public. This study succeeded in developing the evaluation system with Large Language Model integrated into an MOOC platform so that the question generator feature and the student evaluation using AI-generated questions are integrated in the same platform. The evaluation performed in this study compares the models used to generate the question and evaluates the experience of using the developed feature. The two said model Mistral-7B and LLaMa-3 8B, with the result that shows Mistral-7B presents the better output. The developed feature is accessible and easy to understand by its users with more room for improvement for further studies.